LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-234770

Nonlinear Statistical Signal Processing: A Particle Filtering Approach

J.V. Candy

September 19, 2007

Imaging for Detection and Identification

NONLINEAR STATISTICAL SIGNAL PROCESSING: A PARTICLE FILTERING APPROACH

J. V. Candy (tsoftware@aol.com)
University of California, Lawrence Livermore National Lab. & Santa Barbara
Livermore CA 94551
USA

Abstract. A introduction to particle filtering is discussed starting with an overview of Bayesian inference from batch to sequential processors. Once the evolving Bayesian paradigm is established, simulation-based methods using sampling theory and Monte Carlo realizations are discussed. Here the usual limitations of nonlinear approximations and non-gaussian processes prevalent in classical nonlinear processing algorithms (e.g. Kalman filters) are no longer a restriction to perform Bayesian inference. It is shown how the underlying hidden or state variables are easily assimilated into this Bayesian construct. Importance sampling methods are then discussed and shown how they can be extended to sequential solutions implemented using Markovian state-space models as a natural evolution. With this in mind, the idea of a particle filter, which is a discrete representation of a probability distribution, is developed and shown how it can be implemented using sequential importance sampling/resampling methods. Finally, an application is briefly discussed comparing the performance of the particle filter designs with classical nonlinear filter implementations.

KEYWORDS: Particle filtering, Bayesian processing, Bayesian approach, simulation-based sampling, nonlinear signal processing

1 Introduction

In this paper we develop the "Bayesian approach" to signal processing for a variety of useful model sets. It features the next generation of processor that has recently been enabled with the advent of high speed/high throughput computers. The emphasis is on nonlinear/non-gaussian problems, but classical techniques are included as special cases to enable the reader familiar with such methods to draw a parallel between the approaches. The common ground is the model sets. Here the state-space approach is emphasized because of its inherent applicability to a wide variety of problems both linear and nonlinear as well as time invariant and

time-varying including what has become popularly termed "physics-based" models. Here we discuss the next generation of processors that will clearly dominate the future of model-based signal processing for years to come [1]. This paper discusses a unique perspective of signal processing from the Bayesian viewpoint in contrast to the pure statistical approach. The underlying theme of this paper is the Bayesian approach which is uniformly developed and followed throughout.

2 Bayesian Approach to Signal Processing

In this section we motivate the idea of Bayesian estimation from the purely probabilistic perspective, that is, we do not consider underlying models at all, just densities and distributions. Modern statistical signal processing techniques evolve directly from a Bayesian perspective, that is, they are cast into a probabilistic framework using Bayes' theorem as the fundamental construct. Bayesian techniques are constructed simply around Bayes' theorem. More specifically, the information about the random signal, $x(t)$, required to solve a vast majority of estimation/processing problems is incorporated in the underlying probability distribution generating the process. For instance, the usual signal enhancement problem is concerned with providing the "best" (in some sense) estimate of the signal at time t based on all of the data available at that time. The filtering distribution provides that information directly in terms of its underlying statistics. That is, by calculating the statistics of the process directly from the filtering distribution, the enhanced signal can be extracted using a variety of estimators like *maximum a posteriori*, *maximum likelihood*, *minimum mean-squared error* accompanied by a variety of performance statistics such as error covariances and bounds [1], [2].

We cast this discussion into a dynamic variable/parameter structure by defining the "unobserved" signal or equivalently "hidden" variables as the set of N_x-vectors, $\{x(t)\}, t = 0, \cdots, N$. On the other hand, we define the observables or equivalently measurements as the set of N_y-vectors, $\{y(t)\}, t = 0, \cdots, N$ considered to be *conditionally independent* of the signal variables. The goal in recursive Bayesian estimation is to sequentially (in-time) estimate the joint *posterior* distribution, $\Pr(x(0), \cdots, x(N); y(0), \cdots, y(N))$. Once the posterior

is estimated, than many of the interesting statistics characterizing the process under investigation can be exploited to extract meaningful information.

We start by defining two sets of random (vector) processes: $X_t := \{x(0), \cdots, x(t)\}$ and $Y_t := \{y(0), \cdots, y(t)\}$. Here we can consider X_t to be the set of dynamic random variables or parameters of interest and Y_t as the set of measurements or observations of the desired process.[1] In any case we start with Bayes' theorem for the joint posterior distribution as

$$\Pr(X_t|Y_t) = \frac{\Pr(Y_t|X_t) \times \Pr(X_t)}{\Pr(Y_t)} \tag{1}$$

In Bayesian theory, the *posterior* defined by $\Pr(X_t|Y_t)$ is decomposed in terms of the *prior* $\Pr(X_t)$, its *likelihood* $\Pr(Y_t|X_t)$ and the *evidence* or normalizing factor, $\Pr(Y_t)$. Each has a particular significance in this construct.

It has been shown [3] the joint posterior distribution can be expressed, sequentially, as the joint *sequential Bayesian posterior estimator* as

$$\Pr(X_t|Y_t) = \left[\frac{\Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1))}{\Pr(y(t)|Y_{t-1})}\right] \Pr(X_{t-1}|Y_{t-1}) \tag{2}$$

This result is satisfying in the sense that we need only know the *joint* posterior distribution at the previous stage, $t-1$, scaled by a weighting function to sequentially propagate the posterior to the next stage, that is,

$$\overbrace{\Pr(X_t|Y_t)}^{NEW} = \overbrace{\mathcal{W}(t,t-1)}^{WEIGHT} \times \overbrace{\Pr(X_{t-1}|Y_{t-1})}^{OLD} \tag{3}$$

where the *weight* is defined by

$$\mathcal{W}(t,t-1) := \left[\frac{\Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1))}{\Pr(y(t)|Y_{t-1})}\right]$$

Even though this expression provides the full joint posterior solution, it is not physically realizable unless the distributions are known in closed form and the underlying multiple integrals or sums

[1] In *Kalman filtering theory*, the X_t are considered the states or hidden variables not necessarily observable directly, while the Y_t are observed or measured directly.

Table 1: Sequential Bayesian Processor for Filtering Posterior

Prediction

$$\Pr(x(t)|Y_{t-1}) = \int \Pr(x(t)|x(t-1)) \times \Pr(x(t-1)|Y_{t-1}) dx(t-1)$$

Update/Posterior

$$\Pr(x(t)|Y_t) = \Pr(y(t)|x(t)) \times \Pr(x(t)|Y_{t-1}) \Big/ \Pr(y(t)|Y_{t-1})$$

Initial Conditions

$$\overline{x}(0) \quad \overline{P}(0) \quad \Pr(x(0)|Y_0)$$

can be analytically determined. In fact, a more useful solution is the *marginal* posterior distribution [3] given by the *update recursion*[2] as

$$\overbrace{\Pr(x(t)|Y_t)}^{Posterior} = \frac{\overbrace{\Pr(y(t)|x(t))}^{Likelihood} \times \overbrace{\Pr(x(t)|Y_{t-1})}^{Prior}}{\underbrace{\Pr(y(t)|Y_{t-1})}_{Evidence}} \tag{4}$$

where we can consider the update or filtering distribution as a weighting of the prediction distribution as in the full joint case above, that is,

$$\overbrace{\Pr(x(t)|Y_t)}^{UPDATE} = \overbrace{\mathcal{W}_c(t,t-1)}^{WEIGHT} \times \overbrace{\Pr(x(t)|Y_{t-1})}^{PREDICTION} \tag{5}$$

where the *weight* in this case is defined by

$$\mathcal{W}_c(t,t-1) := \frac{\Pr(y(t)|x(t))}{\Pr(y(t)|Y_{t-1})}$$

We summarize the sequential Bayesian processor in Table 1.

These two sequential relations form the theoretical foundation of many of the sequential particle filter designs. Next we consider the idea of Bayesian importance sampling.

[2]Note that this expression precisely satisfies Bayes' rule as illustrated in the equation.

3 Monte Carlo Approach

In signal processing, we are interested in some statistical measure of a random signal or parameter usually expressed in terms of its moments. For example, suppose we have some signal function, say $f(X)$, with respect to some underlying probabilistic distribution, $\Pr(X)$, then a typical measure to seek is its performance "on the average" which is characterized by the expectation

$$E_X\{f(X)\} = \int f(X)\Pr(X)dX \tag{6}$$

Instead of attempting to use numerical integration techniques, stochastic sampling techniques known as *Monte Carlo* (*MC*) *integration* have evolved as an alternative. The *key idea* embedded in the *MC* approach is to represent the required distribution as a set of random *samples* rather than a specific analytic function (e.g. Gaussian). As the number of samples becomes large, they provide an equivalent representation of the distribution enabling moments to be estimated directly.

Monte Carlo integration draws samples from the required distribution and then forms sample averages to approximate the sought after distributions, that is, it maps integrals to discrete sums. Thus, *MC* integration evaluates Eq. 6 by drawing samples, $\{X(i)\}$ from $\Pr(X)$ with "$\longrightarrow$" defined as *drawn from*. Assuming perfect sampling, this produces the estimated or *empirical distribution* given by

$$\hat{\Pr}(X) \approx \frac{1}{N}\sum_{i=1}^{N}\delta\left(X - X(i)\right)$$

which is a probability distribution of mass or weights, $\frac{1}{N}$ and random variable or location $X(i)$. Substituting the empirical distribution into the integral gives

$$E_X\{f(X)\} = \int f(X)\hat{\Pr}(X)dX \approx \frac{1}{N}\sum_{i=1}^{N} f(X(i)) \equiv \overline{f} \tag{7}$$

which follows directly from the sifting property of the delta function. Here $\overline{f}$ is said to be a *Monte Carlo estimate* of $E_X\{f(X)\}$.

A generalization to the *MC* approach is known as *importance sampling* which evolves from:

$$I = \int_X g(x)dx = \int_X \left(\frac{g(x)}{q(x)}\right) \times q(x)\ dx \quad \text{for } \int q(x)dx = 1 \qquad (8)$$

Here $q(x)$ is referred to as the sampling distribution or more appropriately the *importance sampling distribution*, since it samples the *target distribution*, $g(x)$, non-uniformly giving "more importance" to some values of $g(x)$ than others. We say that the *support* of $q(x)$ covers that of $g(x)$, that is, the samples drawn from $q(\cdot)$ overlap the same region (or more) corresponding to the samples of $g(\cdot)$. The integral in Eq. 8 can be estimated by:

- Draw N-samples from

$$q(x) : X(i) \longrightarrow q(x) \text{ and } \hat{q}(x) \approx \frac{1}{N}\sum_{i=1}^{N} \delta(x - X(i));$$

- Compute the sample mean,

$$I = E_q\left\{\frac{g(x)}{q(x)}\right\} \approx \int \left(\frac{g(x)}{q(x)}\right) \times \frac{1}{N}\sum_{i=1}^{N} \delta(x - X(i))\ dx = \frac{1}{N}\sum_{i=1}^{N} \frac{g(X(i))}{q(X(i))}$$

Consider the case where we would like to estimate the expectation of the function of X given by $f(X)$, then choosing an importance distribution, $q(x)$, that is similar to $f(x)$ with covering support gives the expectation estimator

$$E_p\{f(x)\} = \int_X f(x) \times p(x)dx = \int_X f(x) \left(\frac{p(x)}{q(x)}\right) \times q(x)dx \qquad (9)$$

If we draw samples, $\{X(i)\},\ i = 0, 1, \cdots, N$ from the importance distribution, $q(x)$, and compute the sample mean, then we obtain the importance sampling estimator

$$E_p\{f(x)\} = \int_X f(x) \left(\frac{p(x)}{q(x)}\right) \times q(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} f(X(i)) \times \left(\frac{p(X(i))}{q(X(i))}\right) \qquad (10)$$

demonstrating the concept. Note we are again assuming perfect (uniform) sampling with $\hat{q}(x) \approx \frac{1}{N}\sum_{i=1}^{N} \delta(x - X(i))$.

The "art" in importance sampling is in choosing the importance distribution, $q(\cdot)$, that approximates the target distribution, $p(\cdot)$, as

closely as possible. This is the *principal factor* effecting performance of this approach, since variates must be drawn from $q(x)$ that cover the target distribution. Using the concepts of importance sampling, we can approximate the posterior distribution with a function on a finite discrete support. Since it is usually not possible to sample directly from the posterior, we use *importance sampling* coupled with an easy to sample proposal distribution, say $q(X_t|Y_t)$—this is the crucial choice and design step required in Bayesian importance sampling methodology. Here $X_t = \{x(0), \cdots, x(t)\}$ represents the set of dynamic variables and $Y_t = \{y(0), \cdots, y(t)\}$, the set of measured data as before. Therefore, starting with a function of the set of variables, say $g(X_t)$, we would like to estimate its mean using the importance concept, that is,

$$E\{g(X_t)\} = \int g(X_t) \times \Pr(X_t|Y_t)\ dX_t \tag{11}$$

where $\Pr(X_t|Y_t)$ is the posterior distribution. Using the *MC* approach, we would like to sample from this posterior directly and then use sample statistics to perform the estimation. Therefore we insert the proposal importance distribution, $q(X_t|Y_t)$ as before

$$\hat{g}(t) := E\{g(X_t)\} = \int g(X_t) \left[\frac{\Pr(X_t|Y_t)}{q(X_t|Y_t)}\right] \times q(X_t|Y_t)\ dX_t \tag{12}$$

Now applying Bayes' rule to the posterior distribution, and defining an *unnormalized* weighting function as

$$W(t) := \frac{\Pr(X_t|Y_t)}{q(X_t|Y_t)} = \frac{\Pr(Y_t|X_t) \times \Pr(X_t)}{q(X_t|Y_t)} \tag{13}$$

and substituting gives

$$\hat{g}(t) = \int \left[\frac{W(t)}{\Pr(Y_t)}\right] g(X_t) \times q(X_t|Y_t)\ dX_t \tag{14}$$

The *evidence* or normalizing distribution, $\Pr(Y_t)$, is very difficult to estimate; however, it can be eliminated in this expression by first replacing it by the total probability and inserting the importance distribution to give

$$\hat{g}(t) = \frac{E_q\{W(t) \times g(X_t)\}}{E_q\{W(t)\}} \tag{15}$$

that are just expectations with respect to the proposal importance distribution.

Thus, drawing samples from the proposal $X_t(i) \rightarrow X_t \sim q(X_t|Y_t)$ and using the MC approach (integrals to sums) leads to the desired result. That is, from the "perfect" sampling distribution, we have that

$$\hat{q}(X_t|Y_t) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(X_t - X_t(i)) \tag{16}$$

and therefore substituting, applying the sifting property of the Dirac delta function and defining the "normalized" weights as before by

$$\mathcal{W}_i(t) := \frac{W_i(t)}{\sum_{i=1}^{N} W_i(t)} \quad \text{for} \quad W_i(t) = \frac{\Pr(Y_t|X_t(i)) \times \Pr(X_t(i))}{q(X_t(i)|Y_t)} \tag{17}$$

we obtain the final estimate

$$\hat{g}(t) \approx \sum_{i=1}^{N} \mathcal{W}_i(t) \times g(X_t(i)) \tag{18}$$

This importance estimator is biased being the ratio of two sample estimators, but it can be shown that it asymptotically converges to the true statistic and the central limit theorem holds ([4],[5],[6]). Thus as the number of samples increase ($N \rightarrow \infty$), a reasonable estimate of the posterior is

$$\hat{\Pr}(X_t|Y_t) \approx \sum_{i=1}^{N} \mathcal{W}_i(t) \times \delta(X_t - X_t(i)) \tag{19}$$

which is the goal of Bayesian estimation. This estimate provides a "batch" solution, but we must develop a sequential estimate from a more pragmatic perspective.

The importance distribution can be modified to enable a sequential estimation of the desired posterior distribution, that is, we estimate the posterior, $\hat{\Pr}(X_{t-1}|Y_{t-1})$ using importance weights, $\mathcal{W}(t-1)$. As a new sample becomes available, we estimate the new weights, $\mathcal{W}(t)$ leading to an updated estimate of the posterior, $\hat{\Pr}(X_t|Y_t)$. This means that in order to obtain the new set of samples, $X_t(i) \sim q(X_t|Y_t)$ sequentially, we must use the previous set of samples, $X_{t-1}(i) \sim q(X_{t-1}|Y_{t-1})$. Thus, with this in mind, the im-

portance distribution, $q(X_t|Y_t)$ must admit a marginal distribution, $q(X_{t-1}|Y_{t-1})$ implying a Bayesian factorization

$$q(X_t|Y_t) = q(X_{t-1}|Y_{t-1}) \times q(x(t)|X_{t-1}, Y_t) \tag{20}$$

This type of importance distribution leads to the desired sequential solution [7].

Recall the Bayesian solution to the batch posterior estimation problem (as before),

$$\Pr(X_t|Y_t) = \left[\frac{\Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1))}{\Pr(y(t)|Y_{t-1})}\right] \times \Pr(X_{t-1}|Y_{t-1})$$

and recognizing the denominator as just the evidence or normalizing distribution and not a function of X_t, we have

$$\Pr(X_t|Y_t) \propto \Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1)) \times \Pr(X_{t-1}|Y_{t-1}) \tag{21}$$

Substituting this expression for the posterior in the weight relation, we have

$$W(t) = \frac{\Pr(Y_t|X_t)}{q(X_t|Y_t)} = \Pr(y(t)|x(t)) \times \frac{\Pr(x(t)|x(t-1))}{q(x(t)|X_{t-1}, Y_t)} \times \underbrace{\frac{\Pr(X_{t-1}|Y_{t-1})}{q(X_{t-1}|Y_{t-1})}}_{Previous\ Weight} \tag{22}$$

which can be written as

$$W(t) = W(t-1) \times \frac{\Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1))}{q(x(t)|X_{t-1}, Y_t)} \tag{23}$$

giving us the desired relationship—a sequential updating of the weight at each time-step. These results then enable us to formulate a generic Bayesian *sequential importance sampling* algorithm:

1. Choose samples from the proposed importance distribution:

$$x_i(t) \sim q(x(t)|X_{t-1}, Y_t);$$

2. Determine the required conditional distributions:

$$\Pr(x_i(t)|x(t-1)), \quad \Pr(y(t)|x_i(t));$$

3. Calculate the unnormalized weights: $W_i(t)$ using Eq. 23 with

$$x(t) \rightarrow x_i(t);$$

4. Normalize the weights: $\mathcal{W}_i(t)$ of Eq. 17; and
5. Estimate the posterior distribution:

$$\hat{\Pr}(X_t|Y_t) = \sum_{i=1}^{N} \mathcal{W}_i(t)\delta\left(x(t) - x_i(t)\right)$$

Once the posterior is estimated, then desired statistics evolve directly. Next we consider using a model-based approach incorporating state-space models [1].

4 Bayesian Approach to the State-Space

Bayesian estimation relative to the state-space models is based on extracting the unobserved or hidden dynamic (state) variables from noisy measurement data. The Markovian state vector with initial distribution, $\Pr(x(0))$, propagates temporally throughout the state-space according to the *probabilistic transition distribution*, $\Pr(x(t)|x(t-1))$, while the conditionally independent measurements evolve from the *likelihood distribution*, $\Pr(y(t)|x(t))$. We see that the dynamic state variable at time t is obtained through the transition probability based on the previous state (Markovian property), $x(t-1)$, and the knowledge of the underlying conditional probability. Once propagated to time t, the dynamic state variable is used to update or correct based on the likelihood probability and the new measurement, $y(t)$. This evolutionary process is illustrated in Fig. 1. Note that it is the knowledge of these conditional distributions that enable the Bayesian processing.

The usual model-based constructs of the dynamic state variables indicate that there is an equivalence between the probabilistic distributions and the underlying state/measurement transition models. The functional discrete state representation of the previous section given by

$$\begin{aligned} x(t) &= \mathcal{A}\left(x(t-1), u(t-1), w(t-1)\right) \\ y(t) &= \mathcal{C}\left(x(t), u(t), v(t)\right) \end{aligned} \tag{24}$$

where w and v are the respective process and measurement noise sources with u a known input. Here $\mathcal{A}(\cdot)$ is the nonlinear (or linear) dynamic state transition function and $\mathcal{C}(\cdot)$ the corresponding measurement function. Both conditional probabilistic distributions embedded within the Bayesian framework are *completely* specified by these functions and the underlying noise distributions: $\Pr(w(t-1))$ and $\Pr(v(t))$. That is, we have the equivalence[3]

$$\begin{aligned} \mathcal{A}(x(t-1), u(t-1), w(t-1)) &\Rightarrow \Pr(x(t)|x(t-1)) \Leftrightarrow \mathcal{A}(x(t)|x(t-1)) \\ \mathcal{C}(x(t), u(t), v(t)) &\Rightarrow \Pr(y(t)|x(t)) \quad \Leftrightarrow \mathcal{C}(y(t)|x(t)) \end{aligned} \tag{25}$$

Thus, the state-space model along with the noise statistics and prior distributions define the required Bayesian representation or probabilistic propagation model defining the evolution of the states and measurements through the transition probabilities. This is a subtle point that must be emphasized and illustrated in the diagram of Fig. 1. Here the dynamic state variables propagate throughout the state-space specified by the transition probability ($\mathcal{A}(x(t)|x(t-1))$) using the embedded process model. That is, the "unobserved" state at time $t-1$ depends the transition probability distribution to propagate to the state at time t. Once evolved, the state combines under the corresponding measurement at time t through the conditional likelihood distribution ($\mathcal{C}(y(t)|x(t))$) using the embedded measurement model to obtain the required likelihood distribution. These events continue to evolve throughout with the states propagating through the state transition probability using process model and the measurements generated by the states and likelihood using the measurement model. From the Bayesian perspective, the broad initial prior is scaled by the evidence and "narrowed" by the likelihood to estimate the posterior.

With this in mind we can now return to the original Bayesian estimation problem, define it and show (at least conceptually) the solution. Using the state-space and measurement representation, the basic dynamic state estimation (signal enhancement) problem can now be stated in the Bayesian framework as:

[3]We use this notation to emphasize the influence of both process ($\mathcal{A}$) and measurement ($\mathcal{C}$) representation on the conditional distributions.

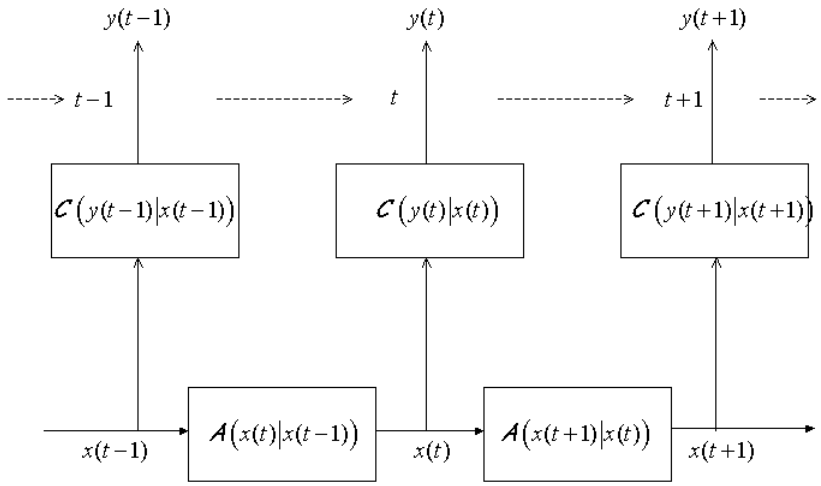

Figure 1: Bayesian State-Space Probabilistic Evolution.

GIVEN a set of noisy uncertain measurements, $\{y(t)\}$, and known inputs, $\{u(t)\}$;$t = 0, \cdots, N$ along with the corresponding prior distributions for the initial state and process and measurement noise sources: $\Pr(x(0))$, $\Pr(w(t-1))$, $\Pr(v(t))$ as well as the conditional transition and likelihood probability distributions: $\Pr(x(t)|x(t-1))$, $\Pr(y(t)|x(t))$ characterized by the state and measurement models: $\mathcal{A}(x(t)|x(t-1))$, $\mathcal{C}(y(t)|x(t))$, FIND the "best" (filtered) estimate of the state, $x(t)$, say $\hat{x}(t|t)$ based on all of the data up to and including t, Y_t, that is, find the best estimate of the filtering *posterior*, $\Pr(x(t)|Y_t)$, and its associated statistics.

Analytically, to generate the model-based version of the sequential Bayesian processor, we replace the transition and likelihood distributions with the conditionals of Eq. 25. The solution to the signal enhancement or equivalently state estimation problem is given by the filtering distribution, $\Pr(x(t)|Y_t)$ which was solved previously in Sec. 2 (see Table 1). We start with the prediction recursion characterized by the *Chapman-Kolmogorov equation* replacing transition probability with the implied model-based conditional, that is,

$$\Pr(x(t)|Y_{t-1}) = \int \overbrace{\mathcal{A}(x(t)|x(t-1))}^{\text{Embedded Process Model}} \times \overbrace{\Pr(x(t-1)|Y_{t-1})}^{\text{Prior}} dx(t-1) \tag{26}$$

Next we incorporate the model-based likelihood into the posterior equation with the understanding that the process model has been incorporated into the prediction

$$\Pr(x(t)|Y_t) = \overbrace{\mathcal{C}(y(t)|x(t))}^{\text{Embedded Measurement Model}} \times \overbrace{\Pr(x(t)|Y_{t-1})}^{\text{Prediction}} \Big/ \Pr(y(t)|Y_{t-1}) \tag{27}$$

Thus, we see from the Bayesian perspective that the sequential Bayesian processor employing the state-space representation of Eq. 24 is straightforward. Next let us investigate a more detailed development of the processor resulting in a closed-form solution—the linear Kalman filter.

5 Bayesian Particle Filters

Particle filtering (*PF*) is a sequential Monte Carlo method employing the recursive estimation of relevant probability distributions using the concepts of "importance sampling" and the approximations of distributions with discrete random measures ([7]-[11]). The key idea is to represent the required posterior distribution by a set of N_p-random samples, the *particles*, with associated weights, $\{x_i(t), \mathcal{W}_i(t)\}; i = 1, \cdots, N_p$, and compute the required Monte Carlo estimates. Of course, as the number of samples become very large the *MC* representation becomes an equivalent characterization of the analytical description of the *posterior* distribution.

Thus, particle filtering is a technique to implement recursive Bayesian estimators by *MC* simulation. It is an alternative to approximate Kalman filtering for nonlinear problems ([1],[2],[7]). In *PF* continuous distributions are approximated by "discrete" random measures composed of these weighted particles or point masses where the *particles* are actually samples of the unknown or hidden states from the state-space and the *weights* are the associated "probability masses" estimated using the Bayesian recursions as shown in

Fig. 2. From the figure we see that associated with each particle, $x_i(t)$ is a corresponding weight or (probability) mass, $\mathcal{W}_i(t)$. Therefore knowledge of this random measure, $\{x_i(t), \mathcal{W}_i(t)\}$ characterizes the empirical posterior distribution—an estimate of the filtering posterior, that is,

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t)\delta(x(t) - x_i(t))$$

at a particular instant of time t. Importance sampling plays a crucial role in state-space particle algorithm development. *PF* does *not* involve linearizations around current estimates, but rather approximations of the desired distributions by these discrete measures. In comparison, the Kalman filter, recursively estimates the conditional mean and covariance that can be used to characterize the filtering posterior, $\Pr(x(t)|Y_t)$ under gaussian assumptions [1].

In summary, a particle filters is a sequential *MC* based "point mass" representation of probability distributions. They only require a state-space representation of the underlying process to provide a set of particles that evolve at each time step leading to an *instantaneous* approximation of the target posterior distribution of the state at time t given all of the data up to that time. Fig. 2 illustrates the evolution of the posterior at a particular time step. Here we see the estimated posterior based on 21-particles (non-uniformly spaced) and we select the 5-th particle and weight to illustrate the instantaneous approximation at time t for x_i vs $\hat{\Pr}(x(t)|Y_t)$. Statistics are calculated across the ensemble created over time to provide the inference estimates of the states. For example, the *minimum mean-squared error* (*MMSE*) estimate is easily determined by averaging over $x_i(t)$, since

$$\begin{aligned}
\hat{x}_{\text{mmse}}(t) &= \int x(t)\Pr(x(t)|Y_t)dx \approx \int x(t)\hat{\Pr}(x(t)|Y_t)dx \\
&= \frac{1}{N_p}\sum_{i=1}^{N_p} x(t)\mathcal{W}_i(t)\delta(x(t) - x_i(t)) = \frac{1}{N_p}\sum_{i=1}^{N_p} \mathcal{W}_i(t)x_i(t)
\end{aligned}$$

while the maximum a posterior (*MAP*) estimate is simply determined by finding the sample corresponding to the maximum weight

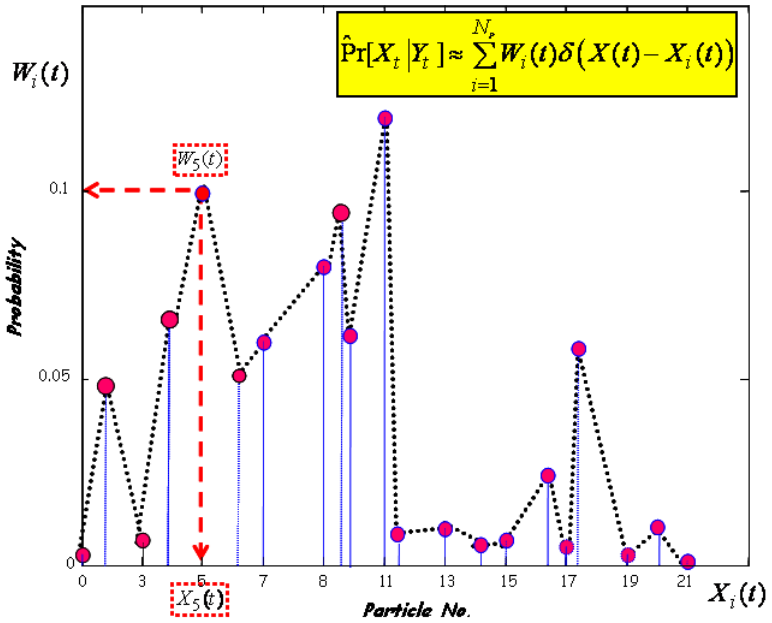

Figure 2: Particle filter representation of posterior probability distribution in terms of weights (probabilities) and particles (samples).

of $x_i(t)$ across the ensemble at each time step, that is,

$$\hat{x}_{\text{MAP}}(t) = \max_{x_i} \left\{ \hat{\Pr}(x(t)|Y_t) \right\} \tag{28}$$

The sequential importance sampling solution to the recursive Bayesian state estimation problem was given previously starting with the recursive form for the importance distribution as

$$q(X_t|Y_t) = q(X_{t-1}|Y_{t-1}) \times q(x(t)|X_{t-1}, Y_t)$$

and evolving to the recursive expression for the importance weights as

$$W(t) = \frac{\Pr(X_t|Y_t)}{q(X_t|Y_t)} = \frac{\Pr(Y_t|X_t) \times \Pr(X_t)}{q(X_{t-1}|Y_{t-1}) \times q(x(t)|X_{t-1}, Y_t)}$$

$$W(t) = W(t-1) \times \frac{\overbrace{\Pr(y(t)|x(t))}^{\text{Likelihood}} \times \overbrace{\Pr(x(t)|x(t-1))}^{\text{Transition}}}{q(x(t)|X_{t-1}, Y_t)} \tag{29}$$

The *state-space particle filter* (*SSPF*) evolving from this sequential importance sampling construct follows directly after sampling from the importance distribution, that is,

$$\begin{aligned} x_i(t) &\longrightarrow q(x(t)|x(t-1), y(t)) \\ W_i(t) &= W_i(t-1) \times \frac{\mathcal{C}(y(t)|x_i(t)) \times \mathcal{A}(x_i(t)|x_i(t-1))}{q(x_i(t)|x_i(t-1), y(t))} \\ \mathcal{W}_i(t) &= \frac{W_i(t)}{\sum_{i=1}^{N_p} W_i(t)} \end{aligned} \tag{30}$$

and the filtering *posterior* is estimated by

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t) \times \delta(x(t) - x_i(t)) \tag{31}$$

assuming that

$$q\left(x(t)|X_{t-1}, Y_t\right) \longrightarrow q\left(x(t)|x(t-1), y(t)\right) \tag{32}$$

then the importance distribution is *only* dependent on $[x(t-1), y(t)]$ which is common when performing filtering, $\Pr(x(t)|Y_t)$, at each instant of time.

This completes the theoretical motivation of the state-space-particle filters. Next we consider a pragmatic approach for implementation.

6 Bootstrap Particle Filter

In the previous section, we developed the generic *SSPF* from the simulation-based sampling perspective. The basic design tool when developing these algorithms is the choice of the importance sampling distribution, $q(\cdot)$. One of the most popular realizations of this approach was developed by using the transition prior as the importance proposal [3]. This *prior* is defined in terms of the state-space representation by $\mathcal{A}(x(t)|x(t-1)) \rightarrow \mathcal{A}(x(t-1), u(t-1), w(t-1))$ which is dependent on the known excitation and process noise statistics. It is given by

$$q_{prior}(x(t)|x(t-1), Y_t) \longrightarrow \Pr(x(t)|x(t-1))$$

Substituting this choice into the expression for the weights gives

$$
\begin{aligned}
W_i(t) &= W_i(t-1) \times \frac{\Pr(y(t)|x_i(t)) \times \Pr(x(t)|x_i(t-1))}{q_{prior}(x(t)|x_i(t-1), Y_t)} \\
&= W_i(t-1) \times \Pr(y(t)|x_i(t))
\end{aligned}
$$

since the priors cancel. Note two properties of this choice of importance distribution. First, the weight does *not* use the most recent observation, $y(t)$ and second this choice is easily implemented and updated by simply evaluating the measurement likelihood, $\mathcal{C}(y(t)|x_i(t)); i = 1, \cdots, N_p$ for the sampled particle set. These weights require the particles to be propagated to time t *before* the weights can be calculated.

This choice can lead to problems, since the transition prior is not conditioned on the measurement data, especially the most recent. Failing to incorporate the latest available information from the most recent measurement to propose new values for the states leads to only a few particles have significant weights when their likelihood is calculated. The transitional prior is a much broader distribution than the likelihood indicating that only a few particles will be assigned a large weight. Thus, the algorithm will degenerate rapidly. Thus, the *SSPF* algorithm takes the same generic form as before with the importance weights much simpler to evaluate with this approach. It has been called the *bootstrap PF*, the *condensation PF*, or the survival of the fittest algorithm [3].

One of the major problems with the importance sampling algorithms is the depletion of the particles, that is, they tend to increase in variance at each iteration. The degeneracy of the particle weights creates a problem that must be resolved before these particle algorithms can be of any pragmatic use in applications. The problem occurs because the variance of the importance weights can only increase in time [3] thereby making it impossible to avoid this weight degradation. Degeneracy implies that a large computational effort is devoted to updating particles whose contribution to the posterior is negligible. Thus, there is a need to somehow resolve this problem to make the simulation-based techniques viable. This requirement leads to the idea of "resampling" the particles.

The main objective in simulation-based sampling techniques is to generate *i.i.d.* samples from the targeted posterior distribution in

order to perform statistical inferences extracting the desired information. Thus, the importance weights are quite critical since they contain probabilistic information about each specific particle. In fact, they provide us with information about "how probable a sample has been drawn from the target posterior" [12], [13]. Therefore, the weights can be considered acceptance probabilities enabling us to generate independent (approximately) samples from the posterior, $\Pr(x(t)|Y_t)$. The empirical distribution, $\hat{\Pr}(x(t)|Y_t)$ is defined over a set of finite (N_p) random measures, $\{x_i(t), W_i(t)\}; i = 1, \cdots, N_p$ approximating the posterior, that is,

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} W_i(t)\delta(x(t) - x_i(t)) \tag{33}$$

Resampling, therefore, can be thought of as a realization of enhanced particles, $\hat{x}_k(t)$, extracted from the original samples, $x_i(t)$ based on their "acceptance probability", $W_i(t)$ at time t, that is, statistically we have

$$\Pr\Big(\hat{x}_k(t) = x_i(t)\Big) = W_i(t) \qquad \text{for } i = 1, \cdots, N_p \tag{34}$$

or we write it symbolically as

$$\hat{x}_k(t) \Rightarrow x_i(t)$$

with the set of new particles, $\{\hat{x}_k(t)\}$, replacing the old set, $\{x_i(t)\}$.

The fundamental concept in resampling theory is to preserve particles with large weights (large probabilities) while discarding those with small weights. Two steps must occur to resample effectively: (1) a decision, on a weight-by-weight basis, must be made to *select* the appropriate weights and *reject* the inappropriate; and (2) resampling must be performed to minimize the degeneracy. The overall strategy when coupled with importance sampling is termed sequential importance resampling (*SIR*) [3]. We illustrate the evolution of the particles through a variety of prediction-update time-steps in Fig. 3 where we see evolution of each set of particles through prediction, resampling and updating.

We summarize the *bootstrap particle filter* algorithm in Table 4. This completes the algorithm, next we apply it to a standard problem.

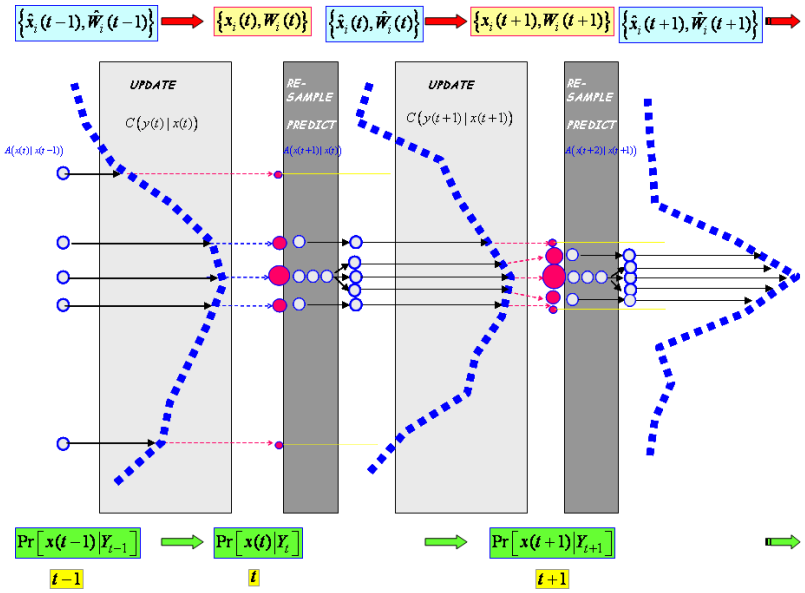

Figure 3: Evolution of particle filter weights and particles using the sequential state-space *SIR* algorithm: resampling, propagation (state-space transition model), update (state-space measurement likelihood), resampling ... .

Table 2: BOOTSTRAP *SIR* State-Space Particle Filtering Algorithm

INITIALIZE:

$$x_i(0) \longrightarrow \Pr(x(0)) \qquad W_i(0) = \frac{1}{N_p} \quad i = 1, \cdots, N_p \qquad \text{[sample]}$$

IMPORTANCE SAMPLING:

$$x_i(t) \sim \mathcal{A}\left(x(t)|x_i(t-1)\right); \; w_i \sim \Pr(w_i(t)) \qquad \text{[state transition]}$$

Weight Update

$$W_i(t) = W_i(t-1) \times \mathcal{C}\left(y(t)|x_i(t)\right) \qquad \text{[weights]}$$

Weight normalization

$$\mathcal{W}_i(t) = \frac{W_i(t)}{\sum_{i=1}^{N_p} W_i(t)}$$

RESAMPLING: $(\hat{x}_i(t) \Rightarrow x_j(t))$

DISTRIBUTION:

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t)\delta(x(t)-\hat{x}_i(t)) \qquad \text{[posterior distribution]}$$

STATE ESTIMATION:

$$\hat{x}(t|t) = E\{x(t)|Y_t\} \approx \frac{1}{N_p}\sum_{i=1}^{N_p} \hat{x}_i(t) \qquad \text{[conditional mean]}$$

$$\hat{X}_{MAP}(t) = \max\ \hat{\Pr}(x(t)|Y_t) \qquad \text{[MAP]}$$

7 Example: Nonlinear Non-gaussian Prediction

We consider a well-known problem that has become a benchmark for many of the *PF* algorithms. It is highly nonlinear, non-gaussian and nonstationary and evolves from studies of population growth [3]. The state transition and corresponding measurement model are given by

$$\begin{aligned} x(t) &= \frac{1}{2}x(t-1) + \frac{25x(t-1)}{1+x^2(t-1)} + 8\cos(1.2(t-1)) + w(t-1) \\ y(t) &= \frac{x^2(t)}{20} + v(t) \end{aligned}$$

where $\triangle t = 1.0$, $w \sim \mathcal{N}(0, 10)$ and $v \sim \mathcal{N}(0, 1)$. The initial state is gaussian distributed with $\overline{x}(0) \sim \mathcal{N}(0.1, 5)$.

In terms of the state-space representation, we have

$$\begin{aligned} a[x(t-1)] &= \frac{1}{2}x(t-10) + \frac{25x(t-1)}{1+x^2(t-1)}) \\ b[u(t-1)] &= 8\cos(1.2(t-1)) \\ c[x(t)] &= \frac{x^2(t)}{20} \end{aligned}$$

In the Bayesian framework, we would like to estimate the instantaneous posterior filtering distribution,

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i\delta(x(t) - x_i(t)) \qquad (35)$$

where the unnormalized importance weight is given by

$$W_i(t) = W_i(t-1) \times \frac{\mathcal{C}\left(y(t)|x(t)\right) \times \mathcal{A}\left(x(t)|x(t-1)\right)}{q(x(t)|X_{t-1}, Y_t)} \tag{36}$$

The *weight recursion* for the bootstrap case is $W_i(t) = W_i(t-1) \times \mathcal{C}\left(y(t)|x(t)\right)$. Therefore, for our problem, the Bayesian processor has its state transition probability given by

$$\Pr\left(x(t)|x(t-1)\right) \longrightarrow \mathcal{A}\left(x(t)|x(t-1)\right) \sim \mathcal{N}\left(x(t) : \mathbf{a}[x(t-1)], \mathbf{R}_{ww}\right) \tag{37}$$

Thus, the *SIR* algorithm becomes:

- Draw samples (particles) from the state transition distribution: $x_i(t) \rightarrow \mathcal{N}\left(x(t) : \mathbf{a}[x(t-1)], \mathbf{R}_{ww}\right)$

$$w_i(t) \rightarrow \Pr(w(t)) \sim \mathcal{N}(0, R_{ww})$$

$$x_i(t) = \frac{1}{2}x_i(t-1) + \frac{25x_i(t-1)}{1+x_i^2(t-1)} + 8\cos\left(1.2(t-1)\right) + w_i(t-1)$$

- Estimate the likelihood, $\mathcal{C}\left(\mathbf{y}(t)|x(t)\right) \rightarrow \mathcal{N}\left(\mathbf{y}(t) : \mathbf{c}[x(t)], \mathbf{R}_{vv}(t)\right)$

$$c[x_i(t)] = \frac{x_i^2(t)}{20}$$

- Update and normalize the weight: $\mathcal{W}_i(t) = W_i(t) / \sum_{i=1}^{N_p} W_i(t)$
- Resample: $\hat{x}_i(t) \Rightarrow x_j(t)$
- Estimate the instantaneous posterior:

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i \delta(x(t) - \hat{x}_i(t))$$

- Estimate the corresponding statistics:

$$\begin{aligned}
\hat{X}_{\text{map}}(t) &= \arg\max \hat{\Pr}(x(t)|Y_t) \\
\hat{X}_{\text{mmse}}(t) &= E\{x(t)|Y_t\} = \sum_{i=1}^{N_p} \hat{x}_i(t) \hat{\Pr}(x(t)|Y_t) \\
\hat{X}_{\text{median}}(t) &= \text{median}\left(\hat{\Pr}(x(t)|Y_t)\right)
\end{aligned}$$

We show the simulated data in Fig. 4. In *a* we see the hidden state and *b* the noisy measurement. The estimate instantaneous posterior distribution surface for the state is shown in Fig. 5*a* while

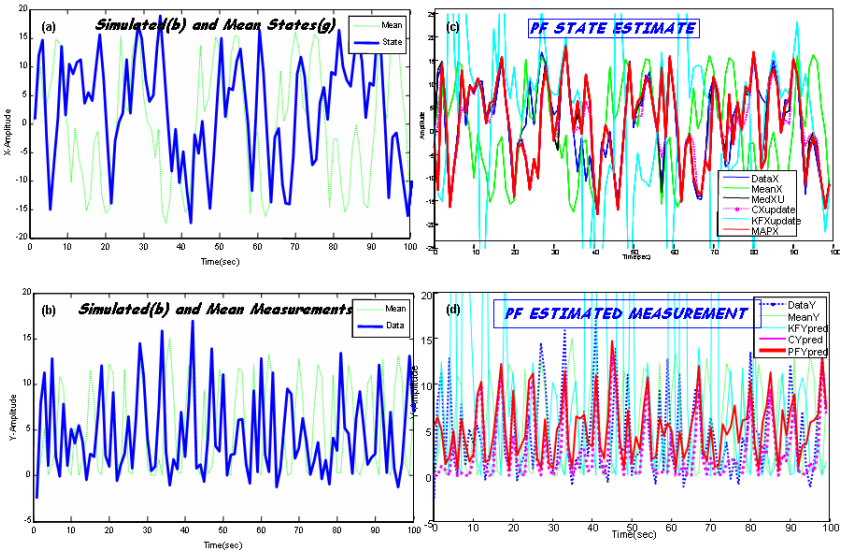

Figure 4: Population growth problem:(a) Simulated state with mean. (b) Simulated measurement with mean. (c) Ensemble of state estimates: median, *EKF MMSE MAP*(d) Ensemble of measurement estimates: median, *EKF MMSE MAP*.

slices at selected instants of time are shown in *b*. Here we see that the posterior is clearly not unimodal and in fact we can see its evolution in time as suggested by Fig. 2 previously. The final state and measurement estimates are shown in Fig. 4*b* demonstrating the effectiveness of the *PF* bootstrap processor for this problem. Various ensemble estimates are shown (e.g. median, *MMSE MAP*). It is clear that the *EKF* gives a poor *MMSE* estimate, since the posterior is *not* gaussian (unimodal).

8 Summary

In this paper we have provided an overview of nonlinear statistical signal processing based on the Bayesian paradigm. We showed that the next generation processors are well-founded on Monte Carlo simulation-based sampling techniques. We reviewed the development of the sequential Bayesian processor using the state-space models. The popular *bootstrap* algorithm was outlined and applied to a standard problem used within the community to test the perfor-

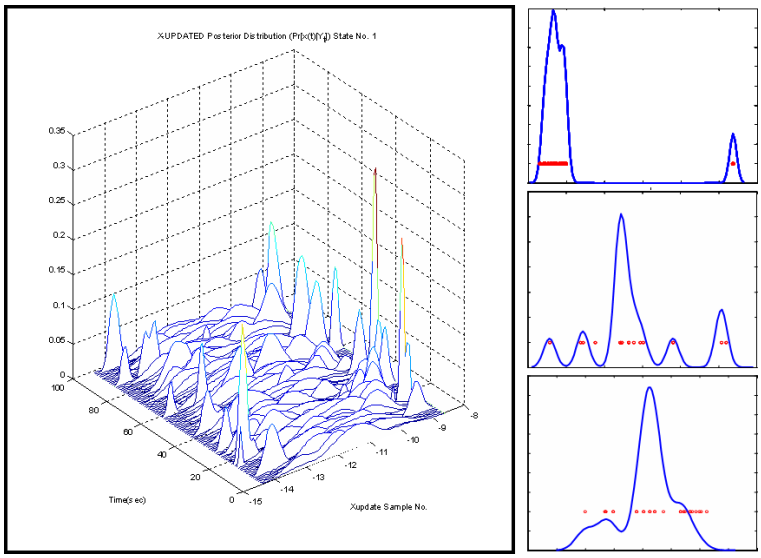

Figure 5: Population growth problem:(a) Instantaneous posterior surface. (b) Time slices of the posterior (cross-section) at selected time-steps.

mance of a variety of *PF* techniques.

References

[1] Candy J. V., *Model-Based Signal Processing*, John Wiley: New Jersey, 2006.

[2] Sullivan E. J. and Candy J. V.,"Space-time array processing: The model-based approach," *J. Acoust. Soc. Am.*, 102, No. 5, 2809-2820 1997.

[3] Doucet A., de Freitas N. and Gordon N., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag: New York, 2001.

[4] Tanner M., *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 2^{nd} *Ed.*, Springer-Verlag: New York,1993.

[5] West M. and Harrison J., *Bayesian Forecasting and Dynamic Models*, 2^{nd} *Ed.*, Springer-Verlag: New York, 1997.

[6] Liu J., *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag: New York, 2001.

[7] Ristic B., Arulampalam S. and Gordon N., *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Boston: Artech House, 2004.

[8] Godsill S. and Djuric P., "Special Issue: Monte Carlo methods for statistical signal processing." *IEEE Trans. Signal Proc.*, vol. 50, 2002.

[9] Djuric P., Kotecha J., Zhang J., Huang Y., Ghirmai T., Bugallo M. and Miguez J., "Particle Filtering." *IEEE Signal Proc. Mag.* vol. 20, No. 5, pp. 19-38, 2003.

[10] Haykin S. and de Freitas N., "Special Issue: Sequential state estimation: from Kalman filters to particle filters." *Proc. IEEE*, vol. 92, No. 3, 2004.

[11] Doucet A. and Wang X., "Monte Carlo methods for signal processing," *IEEE Signal Proc. Mag.* vol. 24, No. 5, pp. 152-170, 2005.

[12] van der Merwe R., *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models* **OGI School of Science & Engr.**, Oregon Health & Science Univ., PhD Dissertation, 2004.

[13] Schoen T., *Estimation of Nonlinear Dynamic Systems: Theory and Applications* **Linkopings Univ.**, Linkoping, Sweden PhD Dissertation, 2006.

This work was performed under the auspices of the U. S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.